Biochem. J. (1985) **229**, 183–188
Printed in Great Britain

Construction and identification of a cDNA clone for human type II procollagen mRNA

Kati ELIMA,* Jyrki K. MÄKELÄ,* Tuula VUORIO,* Sakari KAUPPINEN,†
Jonathan KNOWLES† and Eero VUORIO*
**Department of Medical Chemistry, University of Turku, SF-20520 Turku, Finland and †The State Technical Research Centre of Finland, Biotechnical Laboratory, Recombinant DNA Group, SF-02150 Espoo, Finland*

(*Received 16 October 1984/18 March 1985; accepted 25 March 1985*)

Double-stranded cDNA was constructed for poly(A)-containing RNA isolated from foetal human articular cartilage known to contain small amounts of proα1(II) collagen mRNA. A 585 base pair *Pst*I–*Eco*RI cDNA fragment was isolated and cloned into plasmid pBR322. A resulting recombinant plasmid pHCAR1 was shown to hybridize specifically to a 5.4 kilobase mRNA in cartilage but not in calvarial RNA. Definite identification of clone pHCAR1 was based on sequence analysis; marked homology with the corresponding chick gene and complete agreement with the human gene sequences available were observed.

Type II collagen forms approximately one-half of the organic matrix of articular cartilage. The molecule is synthesized and secreted by chondrocytes into the extracellular space where it plays a role in the differentiation process and participates in the maintenance of tissue strength and flexibility (von der Mark, 1980). At least nine different collagen types are now known. The major collagens expressed during cartilage differentiation and bone formation are of types I and II. Type I collagen consists of two identical α1(I) chains and one α2(I) chain and type II collagen of three α1(II) chains. The different chains are products of different genes. The corresponding mRNAs are translated into (pre)proα chains which undergo extensive post-translational modification, including association of three proα chains into a triple helix, before secretion into the extracellular space for fibril formation (Prockop *et al.*, 1979*a*,*b*).

mRNA from chick cartilage has been used to construct cDNA clones for chick proα1(II) collagen (Vuorio *et al.*, 1982; Ninomiya *et al.*, 1984; Young *et al.*, 1984) with subsequent identification of the gene for chick type II collagen (Sandell *et al.*, 1983, 1984). Chick cDNA clones pCAR1 and pCAR2 (Vuorio *et al.*, 1982) have also been used to isolate genomic clones LgHCol(II)a and LgHCol(II)b for human type II collagen (Strom & Upholt, 1984). The cosmid clone cosH col.1, isolated using chick α1(I) cDNA and previously named human α1(I)-like, has recently been presented as a genomic clone for human α1(II) procollagen (Weiss *et al.*, 1982; Stoker *et al.*, 1984). There is no direct evidence that the corresponding gene is transcribed and processed into mRNA in chondrocytes producing type II collagen. We have recently identified the mRNA for human proα1(II) collagen in total RNA from human foetal cartilage (Vuorio *et al.*, 1984). The poly(A)-containing fraction of this RNA was used to construct cDNA clones for human type II collagen. The present paper describes construction of one such cDNA clone corresponding to exons coding for the carboxyl propeptide region of proα1(II) collagen.

Abbreviations used: bp, base pairs; kb, 1000 bases/base pairs.

Experimental

Materials

Restriction endonucleases, T4 DNA ligase and DNA polymerase I were purchased from Amersham International and Boehringer-Mannheim, reverse transcriptase from Life Sciences, RNAase H and nick translation kits from Bethesda Research Laboratories, [^{32}P]dCTP and the M13 sequencing kit from Amersham International, Sephacryl S-1000 and oligo(dT)–cellulose from Pharmacia, DEAE membrane NA-45 from Schleicher and Schuell and nitrocellulose from Millipore. Phages M13mp10 and M13mp11 were gifts from Dr. P. Mäntsälä (University of Turku).

Purification of RNA

Total RNA was purified from calvaria and articular cartilages of human foetuses obtained at therapeutic abortions, and from calvaria and sterna of 17-day chick embryos as described previously (Rowe *et al.*, 1978). Poly(A)-containing RNA was prepared by three cycles of oligo(dT)–cellulose chromatography (Aviv & Leder, 1972).

Construction of double-stranded cDNA

Unfractionated poly(A)-containing RNA from foetal cartilage (1.3 μg in a 50 μl reaction) was used as template to synthesize single-stranded cDNA as described earlier (Vuorio *et al.*, 1982). The second strand was synthesized essentially as described by Gubler & Hoffman (1983) using RNAase H (25 units/ml) and DNA polymerase I (300 units/ml) but without DNA ligase. The double-stranded cDNA (approx. 100 ng) was digested with *Pst*I and *Eco*RI and electrophoresed on a 1.25% agarose gel. Double-stranded cDNAs with sizes between 520 and 600 bp were collected by binding to a DEAE membrane (Winberg & Hammaskjold, 1980).

Construction of recombinant plasmids

Plasmid pBR322 was digested with *Pst*I and *Eco*RI, electrophoresed on a 0.75% agarose gel followed by isolation of the large fragment by using a DEAE membrane. The vector (100 ng) and the cDNA fraction (approx. 5 ng) were coprecipitated and ligated with T4 DNA ligase (0.5 unit in a 10 μl reaction) at 22°C for 20 h.

Transformation was carried out in 50 mM-$CaCl_2$ as described previously (Mandel & Higa, 1970) using *E. coli* strain DH-1 (a gift from Dr. S. Aho, Harvard University). Forty transformants resistant to tetracycline and sensitive to carbenicillin were obtained.

Colony hybridization

The colonies were grown on Whatman 540 filter paper and amplified by chloramphenicol. The filters were processed and hybridized as described by Thayer (1979) using ^{32}P-labelled purified insert DNA from clone pCAR1 as the probe. After washing, the filters are exposed at −70°C with Kodak X-Omat film using intensifying screens. One clone, pHCAR1, was selected for further characterization.

Plasmid DNA isolation

Plasmid DNA was isolated from *E. coli* by using the cleared lysate method (Kahn *et al.*, 1979) followed by gel filtration on a Sephacryl S-1000 column (Suominen *et al.*, 1984).

RNA gel transfers

Total cellular RNAs were electrophoresed as duplicate samples on 0.75% agarose gels after denaturation with glyoxal and dimethyl sulphoxide (Thomas, 1980). One part of the gel was stained with ethidium bromide to localize the rRNAs; RNA from the other part was transferred by blotting to nitrocellulose. The filters were hybridized with nick translated probes at 44°C for 15–30 h, washed and autoradiographed as described previously (Thomas, 1980).

Hybridization probes

Colony hybridization and RNA transfer filters were hybridized with ^{32}P-labelled purified inserts of pCAR1, a cDNA clone containing sequences complementary to chick proα1(II) collagen mRNA. In addition to pHCAR1, the RNA filters were hybridized with plasmids pHCAL1 [a 670 bp cDNA clone for human proα1(I) collagen mRNA] and pCAL3 [a 506 bp cDNA clone for chick proα2(I) collagen mRNA (J. Mäkelä, T. Vuorio, K. Elima & E. Vuorio, unpublished work)].

Sequencing

Isolated pHCAR1 insert and smaller restriction fragments derived from the insert were recloned in both orientations into M13mp10 and M13mp11 (Messing, 1983) and sequenced by using the method of Sanger *et al.* (1977).

Results and discussion

Construction and screening of the cDNA clones

Total RNA isolated from foetal human articular cartilages has previously been shown to contain proα1(II) collagen mRNA (Vuorio *et al.*, 1984). Due to limited amounts of starting material an attempt was made to find specific restriction fragments in the cDNA for proα1(II) collagen, as in our experience these can be cloned even when present in minute amounts. From the published sequence data on genomic clone LgHCol(II)a, exon 4 was known to contain a *Pst*I site near its 3′-end (Strom & Upholt, 1984). In the genes for human proα1(I), human proα2(I) and chick proα2(I) collagens an *Eco*RI site is identically located in exon 1 close to the termination codon (Bernard *et al.*, 1983*a*,*b*; Fuller & Boedtker, 1981). This *Eco*RI site is also present in clone cosH col.1 as presented recently (Stoker *et al.*, 1984). When combined with the sequence data on chick type II collagen gene (Sandell *et al.*, 1984) the *Pst*I–*Eco*RI cDNA fragment was estimated to have a length of approx. 585 bp. Fragments in this size class were

cloned into pBR322. The 40 recombinant clones obtained were screened for type II sequences in colony hybridization using purified pCAR1 insert as the probe. This clone contains sequences complementary to exons 1–3 in the chick gene. The colony exhibiting the strongest hybridization was selected for further characterization and named pHCAR1.

Tissue specificity of pHCAR1

As type II collagen is found in cartilage but not in calvaria, we performed RNA blot analyses on human and chick RNAs from cartilage and calvaria. Under stringent washing conditions (49°C), nick-translated pHCAR1 was found to hybridize strongly to one mRNA species of 5.4kb present in human cartilage (Fig. 1*a*). Under relaxed washing conditions (44°C) pHCAR1 also hybridized to a somewhat smaller (5.3kb) chick proα1(II) mRNA (Fig. 1*b*). In our previous reports the corresponding chick clones pCAR1 and pCAR2 were shown to hybridize to the same two mRNAs (Vuorio *et al.*, 1982, 1984). The RNA blot filter was also hybridized with ^{32}P-labelled pHCAL1, a cDNA clone for human proα1(I) collagen mRNA. Under relaxed washing conditions this probe hybridized to two mRNA species in human and chick calvaria and slightly to human cartilage RNA, which is known to contain some type I collagen mRNA (Vuorio *et al.*, 1984) (Fig. 1c). Using chick cDNA clone for proα2(I) collagen, pCAL3, hybridization was seen mainly to mRNAs in chick calvaria (Fig. 1*d*). Even long overesposures of the human cartilage RNA blots hybridized with pHCAR1 have not revealed a larger mRNA species which is known to exist for most procollagen mRNAs. The relative amounts of the larger mRNA species have, however, been quite low in all our preparations of foetal human RNA (e.g. Fig. 1*c*).

Physical mapping of pHCAR1

A partial restriction map of pHCAR1 was constructed (Fig. 2*a*).

Sequence analysis

For sequencing, the *Pst*I–*Eco*RI insert was recloned in both orientations in M13mp10 and M13mp11. To obtain the complete sequence of both strands, the *Pvu*II–*Pst*I and *Pvu*II–*Eco*RI fragments were also cloned and sequenced as shown in Fig. 2(*a*). Most of the nucleotide sequence, the derived amino acids and the corresponding chick type II procollagen sequences are shown in Fig. 2(*b*). The amino acids are numbered according to Sandell *et al.* (1984). The whole clone covers the nucleotide sequence from 199 to 787 (with nucleotides 217–757 shown in Fig. 2*b*) in the numbering system of Fuller & Boedtker (1981) for proα1(I) collagen cDNA. The clone contains most

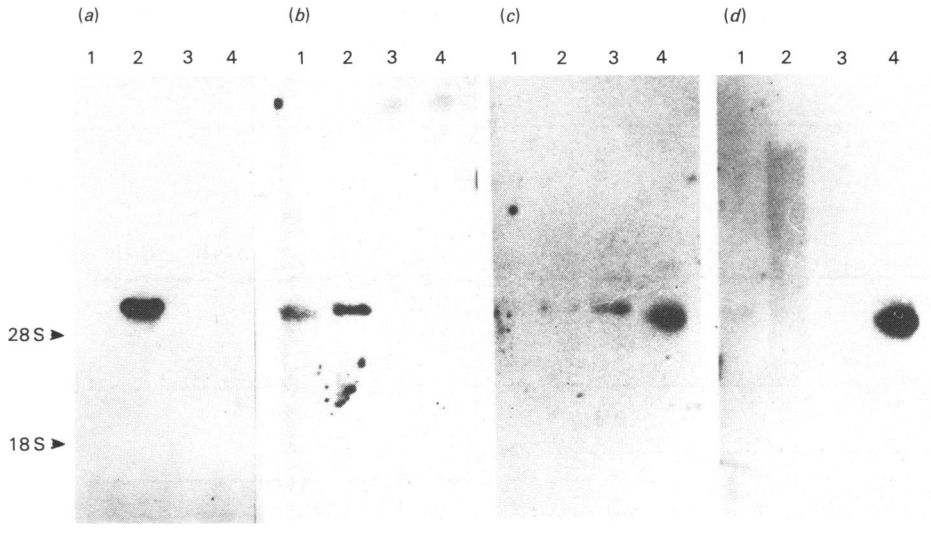

Fig. 1. *RNA blot hybridizations*

Total cellular RNAs (12μg) were denatured with glyoxal, fractionated on a 0.75% agarose gel and blotted to nitrocellulose. Lane 1, embryonic chick sternal RNA; lane 2, foetal human cartilage RNA; lane 3, foetal human calvarial RNA; lane 4, embryonic chick calvarial RNA. The hybridization probes were pHCAR1 (panel *a*, stringent washing; panel *b*, relaxed washing), pHCAL1 (panel *c*) and pCAL3 (panel *d*). The locations of human rRNAs are marked in the Figure.

of the sequence for the non-triple-helical *C*-terminal propeptide and allows for comparisons not only with the corresponding chick sequences but with human proα1(I) and proα2(I) collagen sequences (Bernard *et al.*, 1983*a*,*b*). The human and chick type II collagen sequences show similarities of 83% and 87% at the nucleotide and amino acid levels, respectively. This corresponds well with the homologies between human and chick type I collagen sequencies in this region, which vary between 83% and 88% (Bernard *et al.*, 1983*a*,*b*; Fuller & Boedtker, 1981). Less homology is seen in the nucleotide and amino acid sequences between pHCAR1 and human proα1(I) collagen (75% and 71%) and human proα2(I) collagen (71%, 63%, respectively). Marked homology is also seen in the persistence of the cysteine residues in the human and chick sequences. Deletion of an amino acid at position 126c in chick proα1(II) and proα1(III) propeptides (Yamada *et al.*, 1983) is also seen in human proα1(II), but not in type I procollagen.

Conclusions

This paper describes the construction and identification of the first cDNA clone for human type II procollagen mRNA. The sequence data available on the corresponding gene was used to clone a defined fragment of cDNA. This type of approach is useful in cases when the mRNA is not readily obtainable, to prove that a gene, isolated with the help of a cross-hybridizing probe, is transcribed and expressed. This report provides evidence that the gene corresponding to clones LgHCol(II)a, LgHCol(II)b and cosH col.1 is actively transcribed and processed into mRNA in articular chondrocytes producing type II collagen. Further proof for the identity of clone pHCAR1 is provided by sequence analysis and comparison with the corresponding chick sequences (Fig. 2*b*). No amino acid data on human type II procollagen is available for comparisons with the derived amino acids. The nucleotide sequence of LgH-Col(II)a corresponding to amino acids 73c–79c is identical with pHCAR1 (Strom & Upholt, 1984). Recently sequence covering nearly 5kb of the 3′-end of clone cosH col.1 has become available: the nucleotide sequence of exons 1–4 shows 100% homology with pHCAR1 (Cheah *et al.*, 1985).

Using the genomic clones, polymorphic restriction sites have been detected in human type II collagen gene (Driesel *et al.*, 1982; Pope *et al.*, 1984). These and the availability of both cDNA and genomic clones makes it possible to perform detailed studies on the large number of diseases affecting cartilage development and metabolism.

The expert technical assistance of Merja Haapanen and Petteri Multimäki is gratefully acknowledged. We are grateful to William B. Upholt and Neil G. Stoker for co-operation and valuable advice. This work was supported by grants from the Sigrid Jusélius Foundation and the Medical Research Council of the Academy of Finland.

References

Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 1408–1412

Bernard, M. P., Chu, M.-L., Myers, J. C., Ramirez, F., Eikenberry, E. F. & Prockop, D. J. (1983*a*) *Biochemistry* **22**, 5213–5223

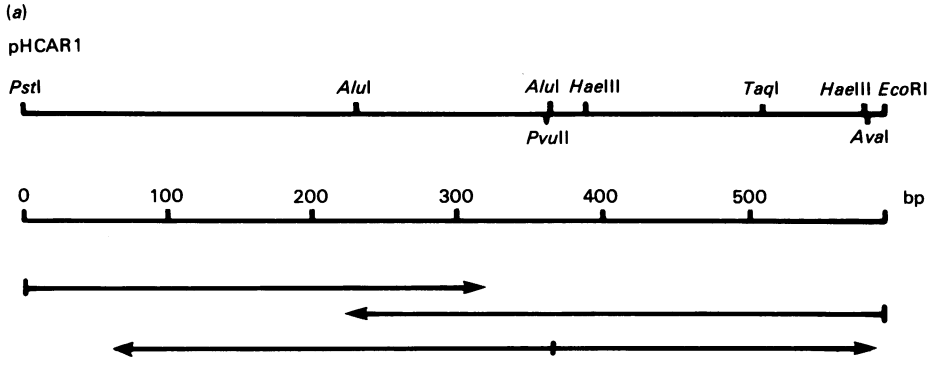

Fig. 2. *Restriction map* (*a*) *and nucleotide sequence* (*b*) *of pHCAR1*
For sequencing restriction fragments were recloned into M13mp10 and M13mp11. Both strands were sequenced as indicated by the arrows under the map. The nucleotide sequence is shown in lane b. The derived amino acid sequence is shown below (lane c). Both sequences are compared with the published nucleotide (lane a) and amino acid (lane d) sequences of chick α1(II) procollagen. Only diverging nucleotides and amino acids are shown in chick sequences. △ indicates the deleted codon and amino acid which is present in human α1(I) procollagen.

(*b*)

```
   73c                          80c
a        T   C                C       T                      G
b  TGC CAC CCT GAG TGG AAG AGT GGA GAC TAC TGG ATT GAC CCC AAC
c  Cys-His-Pro-Glu-Trp-Lys-Ser-Gly-Asp-Tyr-Trp-Ile-Asp-Pro-Asn
d

            90c                                          100c
a    G                              C   A   A
b  CAA GGC TGC ACC TTG GAC GCC ATG AAG GTT TTC TGC AAC ATG GAG
c  Gln-Gly-Cys-Thr-Leu-Asp-Ala-Met-Lys-Val-Phe-Cys-Asn-Met-Glu
d                                Ile

                                 110c
a    A           C                 G  CC   C AGC  G  A C      G
b  ACT GGC GAG ACT TGC GTC TAC CCC AAT CCA GCA AAC GTT CCC AAG
c  Thr-Gly-Glu-Thr-Cys-Val-Tyr-Pro-Asn-Pro-Ala-Asn-Val-Pro-Lys
d                                   Thr     Ser-Ser-Ile     Arg

            120c                                         130c
a                    C            CG        A   C        G     G
b  AAG AAC TGG TGG AGC AGC AAG AGC  Δ  AAG GAG AAG AAA CAC ATC
c  Lys-Asn-Trp-Trp-Ser-Ser-Lys-Ser- Δ  Lys-Glu-Lys-Lys-His-Ile
d                   Thr          Thr          Asp             Val

                                 140c
a           C    G            C   C   T       C           C   C
b  TGG TTT GGA GAA ACC ATC AAT GGT GGC TTC CAT TTC AGC TAT GGA
c  Trp-Phe-Gly-Glu-Thr-Ile-Asn-Gly-Gly-Phe-His-Phe-Ser-Tyr-Gly
d          Ala

            150c                                         160c
a        G   C     T C           C      G  A
b  GAT GAC AAT CTG GCT CCC AAC ACT GCC AAC GTC CAG ATG ACC TTC
c  Asp-Asp-Asn-Leu-Ala-Pro-Asn-Thr-Ala-Asn-Val-Gln-Met-Thr-Phe
d      Glu          Ser                 Ser-Ile

                                 170c
a    G       C           C   G                    G
b  CTA CGC CTG CTG TCC ACG GAA GGC TCC CAG AAC ATC ACC TAC CAC
c  Leu-Arg-Leu-Leu-Ser-Thr-Glu-Gly-Ser-Gln-Asn-Ile-Thr-Tyr-His
d                                              Val

            180c                                         190c
a                     C       C A          G  AG A G           G
b  TGC AAG AAC AGC ATT GCC TAT CTG GAC GAA GCA GCT GGC AAC CTC
c  Cys-Lys-Asn-Ser-Ile-Ala-Tyr-Leu-Asp-Glu-Ala-Ala-Gly-Asn-Leu
d                                Met         Glu-Thr

                                 200c
a        A     A C                 A       C                   A A
b  AAG AAG GCC CTG CTC ATC CAG GGC TCC AAT GAC GTG GAG ATC CGG
c  Lys-Lys-Ala-Leu-Leu-Ile-Gln-Gly-Ser-Asn-Asp-Val-Glu-Ile-Arg
d              Ile

            210c                                         220c
a    C          C                 C       GC  T  T   G     C
b  GCA GAG GGC AAT AGC AGG TTC ACG TAC ACT GCC CTG AAG GAT GGC
c  Ala-Glu-Gly-Asn-Ser-Arg-Phe-Thr-Tyr-Thr-Ala-Leu-Lys-Asp-Gly
d                                   Ser-Val     Glu

                                 230c
a                C   T   C   A                  G   G
b  TGC ACG AAA CAT ACC GGT AAG TGG GGC AAG ACT GTT ATC GAG TAC
c  Cys-Thr-Lys-His-Thr-Gly-Lys-Trp-Gly-Lys-Thr-Val-Ile-Glu-Tyr
d

            240c                                         250c
a        G                 G       G       T G A   T           T
b  CGG TCA CAG AAG ACC TCA CGC CTC CCC ATC ATT GAC ATT GCA CCC
c  Arg-Ser-Gln-Lys-Thr-Ser-Arg-Leu-Pro-Ile-Ile-Asp-Ile-Ala-Pro
d                                      Val
```

Bernard, M. P., Myers, J. C., Chu, M.-L., Ramirez, F., Eikenberry, E. F. & Prockop, D. J. (1983*b*) *Biochemistry* **22**, 1139–1145
Cheah, K. S. E., Stoker, N. G., Griffin, J. R., Grosveld, F. G. & Solomon, E. (1985) *Proc. Natl. Acad. Sci. U.S.A.*, in the press
Driesel, A. J., Schumacher, A. M. & Flavell, R. A. (1982) *Hum. Genet.* **62**, 175–176
Fuller, F. & Boedtker, H. (1981) *Biochemistry* **20**, 996–1006
Gubler, U. & Hoffman, B. J. (1983) *Gene* **25**, 263–269
Kahn, M., Kolter, R., Thomas, C., Figurski, D., Meyer, R., Remant, E. & Helinski, D. R. (1979) *Methods Enzymol.* **68**, 268–280
Mandel, M. & Higa, A. (1979) *J. Mol. Biol.* **53**, 159–162
Messing, J. (1983) *Methods Enzymol.* **101**, 20–78
Ninomiya, Y., Showalter, A. M., van der Rest, M., Seidah, N. G., Chrétien, M. & Olsen, B. R. (1984) *Biochemistry* **23**, 617–624
Pope, F. M., Cheah, K. S. E., Nicholls, A. C., Price, A. B. & Grosveld, F. G. (1984) *Br. Med. J.* **288**, 431–434
Prockop, D. J., Kivirikko, K. I., Tuderman, L. & Guzman, N. A. (1979*a*) *N. Engl. J. Med.* **301**, 13–23
Prockop, D. J., Kivirikko, K. I., Tuderman, L. & Guzman, N. A. (1979*b*) *N. Engl. J. Med.* **301**, 75–85
Rowe, D. W., Moen, R. C., Davidson, J. M., Byers, P. H., Bornstein, P. & Palmiter, R. D. (1978) *Biochemistry* **17**, 1581–1590
Sandell, L. J., Yamada, Y., Dorfman, A. & Upholt, W. B. (1983) *J. Biol. Chem.* **258**, 11617–11621
Sandell, L. J., Prentice, H. L., Kravis, D. & Upholt, W. B. (1984) *J. Biol. Chem.* **259**, 7826–7834
Sanger, F., Nicklen, S. & Coulson, E. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
Stoker, N. G., Cheah, K. S. E., Griffin, J. R., Grosveld, F. G. & Solomon, E. (1984) *J. Cell. Biochem. Suppl.* **8B**, 284
Strom, C. M. & Upholt, W. B. (1984) *Nucleic Acids Res.* **12**, 1025–1038
Suominen, A. I., Karp, M. T. & Mäntsälä, P. I. (1984) *Biochem. Int.* **8**, 209–215
Thayer, R. E. (1979) *Anal. Biochem.* **98**, 60–63
Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 5201–5205
von der Mark, K. (1980) *Curr. Top. Dev. Biol.* **14**, 199–225
Vuorio, E., Sandell, L., Kravis, D., Sheffield, V. C., Vuorio, T., Dorfman, A. & Upholt, W. B. (1982) *Nucleic Acids Res.* **10**, 1175–1192
Vuorio, E., Elima, K., Pulkkinen, J. & Viitanen, A.-M. (1984) *FEBS Lett.* **174**, 238–242
Weiss, E. H., Cheah, K. S. E., Grosveld, F. G., Dahl, H. H. M., Solomon, E. & Flavell, R. A. (1982) *Nucleic Acids Res.* **10**, 1981–1994
Winberg, G. & Hammaskjold, M.-L. (1980) *Nucleic Acids Res.* **8**, 253–264
Yamada, Y., Kühn, K. & de Crombrugghe, B. (1983) *Nucleic Acids Res.* **11**, 2733–2744
Young, M. F., Vogeli, G., Nunez, A. M., Fernandez, M. P., Sullivan, M. & Sobel, M. E. (1984) *Nucleic Acids Res.* **12**, 4207–4228